DOCUMENT RESUME

ED 293 560 IR 052 353

AUTHOR Losee, Robert M., Jr.
TITLE Choosing To Examine Electronic Information: A Message Ranking Model To Minimize Information Overload.
PUB DATE [87]
NOTE 32p.
PUB TYPE Reports - Research/Technical (143) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Costs; *Information Processing; *Information Retrieval; Information Science; *Information Utilization; *Models; *Telecommunications; *User Needs (Information)
IDENTIFIERS *Information Overload; Poisson Probability Distribution

ABSTRACT

Designed to minimize information overload in telecommunications systems, the formal model developed in this paper predicts the usefulness of a message based on the available message features, and may be used to rank messages by expected importance or economic worth. The assumptions of binary and two Poisson independent probabilistic distributions of message feature frequencies are examined, and methods of incorporating these distributions into the ranking model are considered. Ways to incorporate user supplied relevance feedback are suggested, and analytic performance measures are proposed to predict system quality. Problems with this formal model are examined, and areas of future research are suggested. The text is supplemented by two tables and one figure. (33 references). (Author/EW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED293560

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Choosing to Examine Electronic Information: A Message Ranking Model to Minimize Information Overload

Robert M. Losee, Jr.
School of Information and Library Studies
Manning Hall 026A
U. of North Carolina
Chapel Hill, NC 27514 USA

IR052353

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robert M. Losee
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

The decision to examine a message at a particular point in time should be made rationally and economically if the message recipient is to operate efficiently. Electronic mail, electronic bulletin board systems, and telephone systems capable of leaving digitized voice messages can contribute to "information overload." If information overload is the economic loss associated with the examination of a number of non-relevant messages, our model provides a formal method for minimizing expected information overload.

The proposed model predicts the usefulness of a message based on the available message features and may be used to rank messages by expected importance or economic worth. The assumptions of binary and two Poisson independent probabilistic distributions of message feature frequencies are examined, and methods of incorporating these distributions into the ranking model are examined. Ways to incorporate user supplied relevance feedback are suggested. Analytic performance measures are proposed to predict system quality. Problems with the use of this formal model are examined, and areas for future research are suggested.

I. Introduction

With the increase in number and importance of sophisticated telecommunications systems, microcomputers, and computer terminals in offices, we have entered an age where information overload of the individual is an increasingly common occurrence [1]. Information overload has been defined as information received at such a rapid rate that it cannot be assimilated [2]. Hiltz and Turoff suggest that in message systems one may suffer from information entropy, "whereby incoming messages are not sufficiently organized by topic or content to be easily recognized as important" [3]. They further note that "structures that will distinguish communications that are probably of interest from those that probably aren't" are desirable; this is one of the functions of the model proposed below.

We define information overload as the receipt of more information than needed or desired to function effectively and further the goals of an an individual or organization. More formally, receipt of undesirable, or non-relevant messages that result in an economic loss for the recipient represents information overload. A system that explicitly minimizes the probability of retrieval of such economically undesirable messages may be shown to minimize the expected information overload.

We propose a model which determines whether a message should be examined, based on all message features available to

the recipient. Although our model is media independent, we will be particularly concerned with those oral, written, or keyboarded messages in natural languages capable of being stored in digital form and thus easily analyzed by a computer system [4 —— 8]. The model, founded in economic and statistical decision theory [9,10], suggests a method whereby unexamined messages are ordered [11] from those most likely to be of interest, or "relevant" to our needs, vis a vis the organization, to those least likely to be relevant. The recipient of more than one message may examine the message with the greatest expected worth, then the message with the next highest expected worth, and so on until a decision is made that the examination of each unexamined message remaining has a negative expected worth [12,13]. Below, we take advantage of decision theoretic models and assumptions developed for document retrieval and automatic indexing applications [14,15].

II. Message Presentation Systems

We define a message as information made available to an organization or individual who may, in turn, choose to examine, ignore, memorize, or discard the information. Messages may be in any form or transmitted through any medium; the only restriction we will place upon a message is that it have feature values which are available for examination by the message ranking procedure before the decision is made to present or not present the message to the user. For example, the fact that a message did or did not originate at computer "unc.uucp" must be available and single valued if it is to be considered a feature. Models consistent

with fuzzy feature values have been discussed by Bookstein and Ng [16].

Message features may be assigned to a message explicitly or implicitly by either the originator of the message or by the recipient. When formatting an electronic message for transmission, one is often asked for one or more keywords indicative of the subject of the message to be attached to the message header. Other features may not be intentionally assigned by the originator yet would be of significance to the recipient. For example, a message with the explicitly assigned keyword "lunch" would have less significance for me than a message which mentions the name of my favorite Chinese restaurant in the body of the message.

Virtual features assigned values by the recipient may reflect the relationship between the message and the current state of the message presentation system. For example, when reading a bulletin board type of messaging system, such as the USENET news, it may be desirable for the user to be able to examine postings to the network grouped by subject. Messages may thus be assigned the virtual feature <shares feature X with message currently being examined>; the value of this virtual feature will change as different messages are examined.

Existing message-receiver interfaces often have the capability to include or exclude a message based upon one of a small number of features, including <author>, <subject>, <origination date>, etc. [17 —— 20]. Proposed systems often lack the ability to combine features when making decisions. One

exception is the Information Lens system described by Malone et al., which allows rules combining features to be included in the system [21]. An example they provide is

```
  IF      From: Silk, Siegel
THEN      Set Characteristic: VIP

  IF      Message type: Action request [and]
          Characteristic: VIP
THEN      Move to:  Urgent.
```

Messages from Silk or Siegel are labeled as being from a VIP. If an incoming message is from a VIP and requests action, the message should be placed in a location marked "urgent" for immediate examination. This and other recent research have improved the quality of message presentation interfaces, yet formal models for the presentation of messages have received little attention.

III. An Economic Model for Examination Decisions

We now suggest a formal economic rule for deciding whether to examine a message: a message should be selected for examination if the cost of doing so is less than the cost of not doing so. This may be formalized as a decision rule: select a message if and only if $C_s < C_r$, where C_s denotes the cost of selecting a message for examination and C_r the cost of rejecting a message--that is, choosing not to examine a message.

Before deciding to examine a message, these costs must be estimated if the decision is to be rational and economically based. A decision rule based on these expected costs might then be to select a message if and only if the expected cost of

selection for examination is less than the expected cost of non-selection for examination, or,

$$Pr(G|m)C_{S,G} + Pr(B|m)C_{S,B} < Pr(B|m)C_{R,B} + Pr(G|m)C_{R,G},$$

where G and B denote the binary message qualities relevant message and non-relevant message, S and R represent the decision to examine (select) or reject a message, Pr(action|features) denotes the probability of performing action given message features m, and $C_{a,c}$ represents the conceptual cost of performing action a given that condition c holds. If we expect information overload to occur, the expected costs associated with examining a message must exceed the expected costs of non-examination of a message we have decided to examine. If we obey our rule, we should not expect information overload, although some may occur.

Through algebraic manipulation, the following decision rule is obtained: select a message if and only if

$$\frac{Pr(G|m)}{Pr(B|m)} > \frac{C_{S,B} - C_{R,B}}{C_{R,G} - C_{S,G}} = C,$$

where C denotes a cost constant combining our earlier cost figures. Below we denote Pr(G|m)/Pr(B|m) as the Message Status Value of a message with set of features m (MSV_m).

Most message recipients will find it difficult to estimate the cost constant C for a given message, making our model appear difficult to use. We may still use our model by ranking messages by their MSV. We present for user examination the message with the highest MSV, then the message with the second highest MSV, and so on. The selection of messages ceases when the selector feels that an appropriate cutoff point in the

message examination process has been reached. Note that the model does not explicitly claim that the first messages examined are "better" than those examined later in the process. The ranking is performed solely because the cost constant C is unknown. Those messages at the bottom of the ranked list are the messages with MSVs below C whose retrieval would likely result in information overload if given to the user for examination.

Our ranking model requires us to assume that (1) messages are either relevant or non-relevant and (2) costs are identical for all messages involved in the ranking. Both assumptions are inexact, yet they allow the development of a relatively simple model. A decision theoretic model suggesting means to avoid making assumptions of binary quality or identical costs for all messages is suggested by Bookstein [22].

Using Bayesian methods, our MSV may be re-expressed as

$$MSV_m = \frac{Pr(m|G)}{Pr(m|B)} \frac{Pr(G)}{Pr(B)}.$$

Because the prior odds of message relevance, Pr(G)/Pr(B), remain constant for all messages, the expression Pr(G)/Pr(B) may be removed from our definition if the MSV is used only to rank messages for selection. The ranking function may now be written as

$$MSV_m = \frac{Pr(m|G)}{Pr(m|B)}.$$

If the simplifying assumption of feature independence is made, our function may be further modified. The independence assumption states that, if the feature set m is explicitly represented in terms of its components, m_1, m_2, m_3 ...,m_i, then P(m|G)

$= P(m_1|G)P(m_2|G)\ P(m_3|G)...P(m_i|G)$, with a similar expression for non-relevant messages. Use of this simplifying assumption will allow us to develop a more easily understood model but will result in some performance degradation of a system based on the model [23,24]. Similar but computationally more expensive models formally incorporating dependence information are discussed below.

Assuming feature independence, the ranking function becomes

$$\prod_{i=1}^{j} \frac{Pr(m_i|G)}{Pr(m_i|B)},$$

assuming the availability of j features concerning the message.

If the f features with non-zero values are rearranged so they are first and numbered 1 through f, our function becomes

$$\prod_{i=1}^{f} \frac{Pr(m_i|G)}{Pr(m_i|B)} \cdot \frac{Pr(0_i|G)}{Pr(0_i|B)} \cdot \prod_{i=1}^{j} \frac{Pr(0_i|G)}{Pr(0_i|B)}.$$

The right hand portion does not vary among messages, and may be ignored if the function is only used for ranking, suggesting

$$MSV_m = \prod_{i=1}^{f} \frac{Pr(m_i|G)}{Pr(0_i|G)} \Big/ \frac{Pr(m_i|B)}{Pr(0_i|B)}.$$

IV. Binary Independent Features

Those features of a message that take values for m equal to 0 or 1 and are statistically independent of other features are said to be binary independent features. Typical binary information might include whether a message is from one's superior, whether everyone in the organization is scheduled to receive a copy of a message, or whether a message is about a specified topic. It may also be useful to code non-binary data to a binary form. For example, we may categorize all messages received in the past 5 business hours as "recent," and all other messages as "old."

A binary feature's distribution in relevant messages may be described by

$$Pr(m|G,p) = p^m(1-p)^{1-m},$$

with a similar expression for non-relevant messages using the parameter q. Parameter p denotes the probability a feature will have a value of 1 given that a message will be relevant, while q denotes the probability that a feature will have a value 1 given that a message will be non-relevant. Note that subscripts are omitted for notational simplicity where it is obvious that an individual feature is being referred to.

Our criterion function under the binary independence assumption is now

$$\prod_{i=1}^{f} \left(\frac{p_i/(1-p_i)}{q_i/(1-q_i)} \right)^{m_i} .$$

For simplification, this may be transformed to

$$\sum_{i=1}^{f} m_i \log \frac{p_i/(1-p_i)}{q_i/(1-q_i)},$$

if the expression is used only for ranking messages. If a member of the engineering department of a hypothetical organization were to use this model, they would find probabilities and feature relevance discrimination values similar to those in Table 1. Features likely to occur only in relevant messages have a positive discrimination value, those occurring with similar frequencies in both classes of messages have a discrimination value near 0, and those features more likely to occur in non-relevant messages have a negative discrimination value.

We may generalize from this result by noting that if statistical independence of features is assumed under several feature distribution models, we may rank messages for selection by

$$\text{MSV} = \sum_{i=1}^{f} m_i \ \text{Discrimination Value}_i$$

where the Discrimination Value$_i$ component measures the degree to which feature i distinguishes between messages to be examined and messages to be rejected. In the binary independence

model, log((p/(1-p))/(q/(1-q))) represents the discrimination value of a binary independently distributed feature.

V. Poisson Independently Distributed Features

The distribution of occurrences of features that have non-binary integer values may be modeled in some circumstances by the Poisson distribution. Such features may be thought of as having been produced by a conceptual feature generator that produces features at a fixed average rate with the interval between feature occurrences a random exponential variable. The Poisson model has been suggested as providing a satisfactory description of the occurrence frequencies of natural language terms [25 – 27]. Use of this distribution may be helpful if advantage is taken of content analytic methods of counting the number of times particular positive or negative terms or expressions occur in a message [28,29]. Because few short messages have more than a single occurrence of each content bearing word, the Poisson distribution is most effectively used for lengthy messages.

The Poisson distribution suggests that a feature occurs with frequency m with probability

$$Pr(m|G,\lambda) = \exp(-\lambda)\,\lambda^m/m!,$$

LAMBDA

with a similar expression using parameter $\bar{\lambda}$ for non-relevant messages. Parameters λ and $\bar{\lambda}$ represent the average feature frequency of features in relevant and non-relevant messages, respectively.

A MSV based on this distribution is

$$\sum_{i=1}^{f} m_i \log \left(\frac{\lambda_i}{\bar{\lambda}_i} \right),$$

where $\log (\lambda_i / \bar{\lambda}_i)$ represents the discrimination value of feature i.

Feature weights of different distributional types may be summed together if statistical independence of features is assumed. For example, given two features, binary distributed feature 1 and Poisson distributed feature 2, the MSV would be calculated as

$$m_1 \log \left(\frac{p_1/(1-p_1)}{q_1/(1-q_1)} \right) + m_2 \log \left(\frac{\lambda_2}{\bar{\lambda}_2} \right).$$

VI. Binary Dependence Models

Message features are usually statistically dependent and not independent as we have assumed. Several different techniques have been proposed to incorporate dependencies into decision theoretic models similar to the one proposed [30,31]. Such models have met with limited success in other similar applications, due in part to the increased difficulty in estimating the greater number of parameters found when dependence information is incorporated into the model.

The quality of a parameter estimate varies with the number of items used to make the estimate. Thus, in realistic situations, there is a trade off between the system performance obtained by using a more accurate (e.g. dependence) model with less

accurate parameter estimates and performance achieved by providing good estimates of parameters based on large sample sizes with an imprecise model. Dependency models incorporate combinations of features; such combinations occur with less frequency than do the features taken individually, resulting in less accurate estimates of dependent term frequencies. Thus, performance may increase if an unrealistic statistical independence assumption is made (as above) and a small number of parameters are accurately estimated, rather than if a more realistic dependence assumption is made, producing poor estimates for a larger number of parameters [23].

Yu et al. [24] provides a summary of the computational techniques used in models incorporating arbitrary amounts of dependence information. Experimental results suggest that such dependence models may be useful and provide accurate rankings, provided that accurate parameter estimates are available. System tests will be necessary to determine the relative worth of incorporating these computationally expensive dependence assumptions, compared to the independence assumption, into our message ranking model, given the difficulty of estimating parameters accurately.

VII. Relevance Feedback

For our message ranking system to improve its performance over time, it must obtain more accurate estimates of parameter values. This is most easily done through use of relevance feedback, in which the user supplies the system with a binary

relevance judgment about each message. The system then revises its estimates of parameter values and will thus provide more accurate rankings for the user in the future.

In the binary independence model, the parameter p for a given feature may be estimated as

$$p = \frac{S + s}{R + r},$$

where S and R are arbitrarily chosen values such that S < R, s denotes the number of relevant examined messages with the feature, and r denotes the number of relevant messages examined. S and R represent knowledge available prior to examining any relevant messages and s and r represent posterior, or learned knowledge obtained from user supplied relevance feedback judgments. The denominator of our definition may be denoted as R'; it represents our confidence in our current estimate. The effect of examining any additional relevant messages on the estimation of parameters is determined by the relative values of R and the number of messages in the learning sample set. The numerator, denoted as S', may be seen as our best guess of the expected value of p multiplied by the confidence factor R'. If we arbitrarily establish S and R at 1 and 2, respectively, and no relevant messages have been examined, thus making s and r both equal to 0, p would be computed as (1+0)/(2+0) = 1/2. If 17 messages were then examined and found relevant, with 5 having occurrences of the feature in question, our estimate of p would be revised to incorporate the new information: p = (1+5)/(2+17) = 6/19. A similar procedure may be used for estimating the

parameter q for non-relevant messages, with parameters $\bar{S}$, $\bar{R}$, $\bar{s}$, and $\bar{r}$.

Under the two Poisson independence model, we may also incorporate learning into our estimation of the parameter values λ and $\bar{\lambda}$. We define $\lambda = (S+s)/(R+r)$, where S and R are initial values, as with the binary independence model; s denotes the number of times a feature has occurred in the relevant messages examined and r denotes the number of relevant messages examined. A similar definition exists for $\bar{\lambda}$, with parameters $\bar{S}$, $\bar{R}$, $\bar{s}$, and $\bar{r}$.

Accurate estimates of the values of $\bar{S}$ and $\bar{R}$, representing our initial knowledge of the characteristics of non-relevant messages, may be obtained by making a reasonable approximation under either binary or two Poisson independent models [32,33]. We begin by setting $\bar{R}$, representing our degree of confidence in the estimate, to 1. Variable $\bar{S}$ is then estimated as the number of occurrences of a feature in a large sample of messages divided by the number of messages in the sample; we make the assumption that if most messages in an organization are not relevant to a given user, we may estimate the characteristics of the messages a user will find non-relevant as being approximated by the characteristics of all messages.

VIII. An Analytic Performance Model

The performance of the proposed ranking model may be examined probabilistically [13]. We may estimate the probability a message with set of features m is relevant as

$$Pr(G|m) = \frac{Pr(m|G)Pr(G)}{Pr(m)}.$$

Given feature independence, this may be modified to

$$Pr(G|m) = Pr(G) \prod_{i=1}^{f} \frac{Pr(m_i|G)}{Pr(m_i)},$$

where f represents the number of system features. Given historical data, we may directly calculate Pr(G) and the $Pr(m_i)$s, as well as the parameters necessary to compute the feature probabilities $Pr(m_i|G)$.

One use of this model is to estimate the number of relevant messages available. If we want to know the expected number of relevant messages, E(G), from among the first n ranked messages, we calculate

$$E(G) = \sum_{i=1}^{n} Pr(G|m_i),$$

where $Pr(G|m_i)$ represents the probability of relevance for message i.

A second use for our performance model is to aid the user in determining when to cease examining messages [12,13]. Figure 1 provides a graph of the expected performance of the ranking system. At any stage in the message examination process, it is possible to produce such a graph, showing the expected quality of both examined and unexamined messages, given the current estimates of parameters. The display of such a graph on a visual display with an indication of the position of the most recently

examined message allows the user to examine the expected quality of the messages yet unexamined. Such information assists the user in accurately determining when it is appropriate to stop examining messages. Inaccuracies in the graph are due to inaccuracies in parameter estimations. These errors should decrease as more messages are examined and more messages for each relevance class are available for the computation of parameter estimates.

IX. Existing Message Handling Models

We have developed a model which enables us to predict whether a message should be examined, given a set of costs. If the costs are unavailable, the messages may be ranked and the message predicted to be the most useful examined first.

Several traditional message ranking procedures may be described by the decision theoretic model developed above. One traditional means of handing messages is first-come, first-served, where messages are ordered in a First In, First Out (FIFO) queue for examination.

The FIFO model may be incorporated into our model by assigning an ascending message number, such as the time, to each message as it arrives. The discrimination value for the feature <message number> is made a negative value. All other features have zero discrimination value. The weight for message n-1 is thus greater than the weight for message n, which it also predates. Thus message n-1 would be examined before message n, resulting in FIFO examination of messages.

A second traditional model for ranking incoming messages is by category, where all the messages in one category are examined before any messages in a second category are examined, and so on. The categories are each represented by a single feature, such as <message from boss>, <message from spouse>, or <message about widget production>. Messages are assigned to only one category. We may obtain ranking from our economic model consistent with the more traditional model if we renumber our features, for notational simplicity, so that a low numbered feature i, representing category i, is considered to be more important than any higher numbered feature j, j>i. Given an Importance function $I(m_i)$, we see that $I(m_1) > I(m_2) > \ldots I(m_{n-1}) > I(m_n)$ for n features. We now assign ad hoc discrimination values (DV) for each feature in such a way that $DV_1 > DV_2 > \ldots > DV_{n-1} > DV_n$. Given binary independent values for message features and all other features being set to zero, our probabilistic model will result in identical message rankings to those suggested by the more traditional model.

The rules incorporated into the Information Lens system represent a more sophisticated set of capabilities than these traditional message manipulation functions. To obtain rankings equivalent to those provided by the Information Lens system, our decision theoretic model must incorporate dependence information. In an earlier example, we noted a rule stating that if a message arrived from either Silk or Siegel and was an "action request," it was to be treated as urgent. We incorporate this rule into a decision theoretic ranking system by setting appropriate prior

parameters indicating, for example, that we have a great deal of confidence that if a message from Silk is an action request and is received, then it is highly probable that the message is relevant.

X. An Application

Let us hypothesize that we are the director of research and development for a corporation planning to bring widgets to market and wish to use this decision theoretic model to assist in message ranking decisions. We have only two message features available to use in making decisions: (1) whether the message was produced over 5 hours ago, which we assume to be binary independently distributed and (2) the number of times widget occurs in a message, which we assume to be Poisson distributed.

Let us also assume that parameters have been computed for both features. For the binary feature, it was found that in the past, 50 percent of the messages 5 or less hours old were found useful, while only 5 percent of the messages over 5 hours old were found useful, thus $p=.5$ and $q=.05$. For the Poisson distributed feature, it was computed that the term widget occurred an average of two times in useful messages but only an average of 2/3 times in rejected messages, thus $\lambda = 2$ and $\bar{\lambda} = .666$.

Sets of features are given in Table 2 ordered by their MSV values. The computation of the MSV for the highest ranked message in the table is

$$4 \ln \frac{2}{.66} + 1 \ln \frac{.5/(1-.5)}{.05/(1-.05)} = 7.34.$$

Messages would be examined by starting at the highest ranked message and selecting individual messages, moving down the ranked list, until the decision is made to stop.

XI. Summary

We have developed a formal model of the decision process used in deciding whether to examine a message in an organizational setting. The model may be used to rank for examination incoming electronic mail, digitized telephone messages, or electronic publications. Messages may be examined in order of expected worth to the user or in an order suggested by more traditional models, such as FIFO or ranking by membership in one of a number of categories.

A system taking advantage of this cost minimizing model may be formally seen as minimizing the expected information overload, if information overload is defined as the examination of messages with an associated expected loss to the individual or organization. Future researchers may wish to examine those assumptions made explicit by our model. Independence assumptions of the model are unrealistic yet experimental evidence suggests that they may be satisfactory for many purposes. Systems based on independence assumptions may be superior to those assuming statistical dependence because of the difficulty of estimating parameters for large numbers of features. Future use of such models by high volume systems will allow for the determination of

the relative merits of independence assumptions. The Poisson distribution has been shown to be a tractable and reasonable approximation of term distributions in a desired class of messages. Answers to these statistical questions, as well as the development of non-binary cost models of organizational messages, may allow for more accurate and thus more beneficial models of the decision to examine information in organizational settings.

TABLE 1

Sample Parameters and Binary Independent Discrimination Values for an Engineer

Term	p	q	Discrimination Value
welding	.03	.005	1.816
the	.98	.98	0
marketing	.01	.02	-0.703

TABLE 2

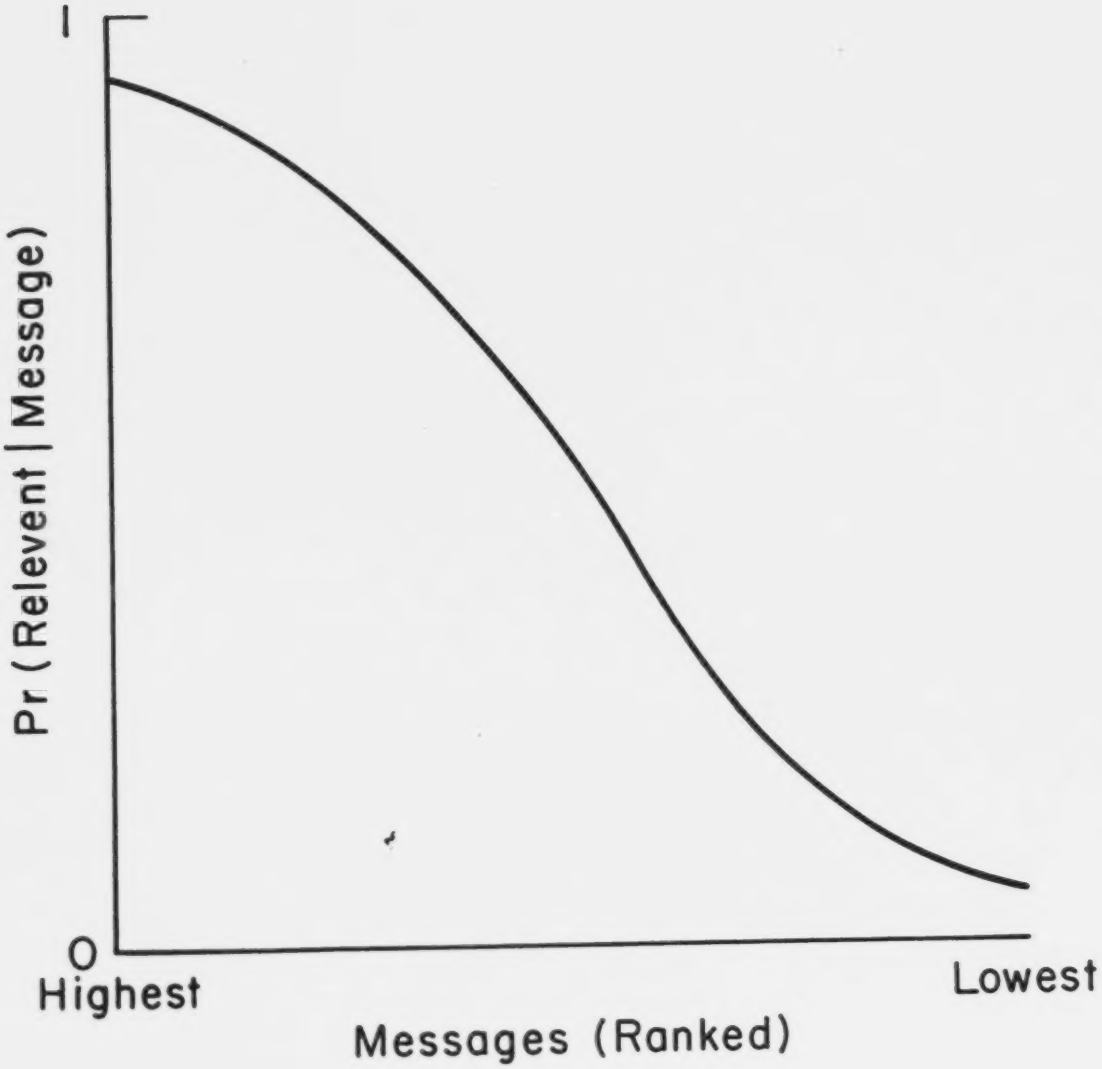
Ranked messages

$\langle m_1, m_2 \rangle$	Message Status Value
<4, 1>	7.34
<4, 0>	4.39
<1, 1>	4.04
<3, 0>	3.30
<0, 1>	2.94
<1, 0>	1.10
<0, 0>	0.00

Feature 1 = frequency of "widget" in a message. p=.5, q=.05
Feature 2 = message over 5 days old (No = 1, Yes = 0) λ_1=2, λ_2=.66.

Figure 1

Expected performance as ranked messages are examined.


1
Pr (Relevent | Message)
0
Highest
Lowest
Messages (Ranked)

References

[1] R. Johansen, J. Vallee, and K. Spangler, Electronic Meetings: Technological Alternatives and Social Choices. Reading, Mass: Addison Wesley, 1979.

[2] T. B. Sheridan and W. R. Ferrell, Man-Machine Systems: Information Control, and Decision Models of Human Performance. Cambridge, Mass: MIT Press, 1974.

[3] S. R. Hiltz and M. Turoff, "Structuring computer-mediated communication systems to avoid information overload." Communications of the ACM, vol 28, pp. 680-689, July 1985.

[4] D. Barcomb, Office Automation, Bedford, Mass: Digital Press, 1981.

[5] C. Ellis and G. Nutt, "Computer science and office information systems," ACM Computing Surveys, vol. 12, pp. 27-60, 1980.

[6] N. B. Finn, The Electronic Office, Prentice-Hall, 1983.

[7] D. Tsichritzis, "Form management," Communications of the ACM, vol. 25, pp. 453-478, 1982.

[8] D. Tsichritzis, ed., Office Automation, Berlin: Springer Verlag, 1985.

[9] J. O. Berger, Statistical Decision Theory, New York: Springer-Verlag, 1980.

[10] R. L. Winkler, Introduction to Bayesian Inference and Decision, New York: Holt, Rinehart, and Winston, 1972.

[11] S. E. Robertson, "The probability ranking principle in IR," Journal of Documentation, vol. 33, pp. 294-304, 1977.

[12] D. H. Kraft and T. Lee, "Stopping rules and their effect on expected search length," Information Processing and Management, vol. 15, pp. 47-58, 1979.

[13] R. M. Losee, "Predicting document retrieval system performance using an expected precision measure," Information Processing and Management, In Press.

[14] C. Van Rijsbergen, Information Retrieval (2nd ed.), London: Butterworths, 1979.

[15] A. Bookstein, "Information retrieval: A sequential learning process," Journal of the American Society for Information Science, vol. 34, pp. 331-342, 1983.

[16] A. Bookstein and K. K. Ng, "A parametric fuzzy set prediction model," Fuzzy Sets and Systems, vol. 17, pp. 131-141, 1985.

[17] P. Denning, "Electronic junk," Communications of the ACM, vol 23, pp. 163-165, 1982.

[18] H. T. Smith, ed., Computer-Based Message Systems, North-Holland, 1984.

[19] M. Stefik, G. Foster, D. Bobrow, K. Kahn, S. Lanning, and L. Suchman, "Beyond the chalkboard: Computer support for collaboration and problem solving in meetings," Communications of the ACM, vol. 30, pp. 32-47, 1987.

[20] E. B. Kerr and S. R. Hiltz, Computer Mediated Communication Systems, New York: Academic Press, 1982.

[21] T. W. Malone, K. R. Grant, F. A. Turbak, S. A. Brobst, and M. D. Cohen, "Intelligent information-sharing systems," Communications of the ACM, vol. 30, pp. 390-402, 1987.

[22] A. Bookstein, "Outline of a general probabilistic retrieval model," Journal of Documentation, vol. 39, pp. 63-72, 1983.

[23] R. M. Losee, A. Bookstein, and C. T. Yu, "Probabilistic models for document retrieval: A comparison of performance on experimental and synthetic databases," Proceedings of the Ninth Annual International ACM SIGIR Conference on Research and Development in Information and Retrieval, (Pisa, Italy: ACM SIGIR), 1986.

[24] C. T. Yu, C. Buckley, K. Lam, and G. Salton, "A generalized term dependence model in information retrieval," Information Technology: Research and Development, vol. 2, pp. 129-54, 1983.

[25] D. C. Stone and M. Rubinoff, "Statistical generation of a technical vocabulary," American Documentation, vol. 19, pp. 411-412, 1968.

[26] S. Harter, A Probabilistic Approach to Keyword Indexing, Unpublished doctoral dissertation, University of Chicago, 1974.

[27] H. P. Luhn, "The automatic creation of literature abstracts," IBM Journal of Research and Development, vol. 2, pp. 159-165, 1958.

[28] D. S. Champion and M. F. Morris, " A content analysis of book reviews in the AJS, ASR, and Social Forces," American Journal of Sociology, vol. 78, pp. 1256-1265, 1973.

[29] K. Krippendorff, Content Analysis: An Introduction to Its Methodology, Beverly Hills: Sage Publications, 1980.

[30] K. Lam, and C. T. Yu, "A clustered search algorithm incorporating arbitrary term dependencies," ACM Transactions on Database Systems, vol. 7, pp. 500-508, 1982.

[31] C. Van Rijsbergen, "A theoretical basis for the use of co-occurrence data in information retrieval," Journal of Documentation, vol. 33, pp. 106-119, 1977.

[32] W. B. Croft and D. Harper, "Using probabilistic models of document retrieval without relevance information," Journal of Documentation, vol. 35, pp. 285-295, 1979.

[33] R. M. Losee, "Parameter estimation for probabilistic document retrieval models," Journal of the American Society for Information Science, In Press.